MONITORING AND IMPROVING THE PERFORMANCE OF STUDENTS BY SENDING ACADEMIC INFORMATION USING IMAP AND ID3

Sateesh kumar Metti[1], and R Raja Ramesh Merugu[2]

[1] M.Tech Student ,Sri Vasavi engineering College T.P Gudem,W.G(D),Andhra Pradesh

[2] *Associate professor,Sri Vasavi Engineering College,T.P Gudem,W.G(D),Andhra Pradesh.*

ABSTRACT:

Monitoring and improving the performance of students is serious issue to be considered in the current education scenario. The main objective of higher education institutions is to provide quality education to its students. The existing solution sends only learning pills to the mobiles using Bluetooth which is confined to those within the range of Bluetooth. In this project this pitfall has been overcome by sending academic information to the mails and sending corresponding alerts to the mobiles. Academic information includes detailed information about the topics covered in the daily classes, Exercises if any, Previous question papers, Mid marks, Class test results and Weekly attendance in a particular course. Which helps us in increasing the motivation and performance of a student and it develops a good understanding of how well or how poorly a student will perform. So, one can take proactive measures to improve the student learning. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. By using ID3 algorithm we extract knowledge that describes students' performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling. The classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here.

KEYWORDS: IMAP, SMTP, Decision Tree, ID3, Prediction Performance

1. INTRODUCTION: Engineering studies are challenged by the proliferation of information, multidisciplinary technological development, globalization, social responsibility, participatory corporate structures, and rapid Change [1]. Engineering students need to develop competencies having three main components:[2] knowledge, the skills they use in managing and applying that knowledge and their attitudes that dictate to which goals their knowledge and skills will be directed. Using emails to create pervasive learning systems [2] and provide contextualized information to students [3], if used in properly designed environments, can promote the better assimilation of knowledge and development of skills[4]. By facilitating the interaction between teachers and students, supporting awareness and understanding of the student context, and intuitively supporting proactively and availability [5]. This paper presents the results of the deployment of a proactive and contextualized system that sends concise exercises, materials, attendance, class test grade, previous question papers, and mid marks to the students' E-mail and corresponding alerts to mobile devices. Data mining provides many tasks that could be used to study the student performance [11]. In this research, the classification task is used to evaluate student's performance tasks that could be used to study the student performance. In this research, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test grade, mid marks and was collected from the student's management system, to predict the performance at the end of the semester. This paper investigates the accuracy of Decision tree techniques for predicting student performance. This data promote reflection and self-study.

2. PROTOCOLS AND DECISION TREE METHODS:

Protocol: A standard set of regulations and requirements that allow two electronic items to connect to and exchange information with one another. In particular, protocols decide [11]: the method of error checking, how to compact data (if required), how the transmitting device signals that it has concluded sending data, and how the receiving device signals that it has completed receiving data. Some examples of these different protocols like SSL, TCP/IP, PPP,SMTP,IMAP,POP, HTTP, and FTP[15].

Here we are using IMAP (Internet Message Access Protocol) [16] and SMTP(Simple Mail Transfer Protocol).IMAP is another most prevalent protocol of internet standard for email usage apart from POP[6]. Usually all the modern email server and client supports these two protocols for transmitting the email messages [16]. For Example Gmail server uses to transmit the message to a client such as Mozilla Thunderbird and Microsoft Outlook.

SMTP [15] is for sending email between 'servers'. Most of the emailing systems implement the messages over internet use SMTP. The message sent from one server to another server, and then the message can be retrieved by an email client.

Data mining : In general Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Technically,

3352

data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [11].

A. Classification: Classification is the most commonly applied data mining technique, which employs a set of pre classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network based classification algorithms. The data classification process involves learning and classification [11].

B. Clustering: Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes [11].

C. Predication: Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables [11].

D. Association rule: Association and correlation is usually to find frequent item [11] set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one [11].

E. Decision Trees: Decision tree is tree shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID) [11].

3. USING E-MAILS AND MOBILE DEVICES FOR MOTIVATING STUDENTS:

The mobile [14] & E-mail learning movement is beginning to gain momentum. Teaching in universities, colleges (and even learning from home) has the potential to be more engaging with the integration of E-mail [6] & m-learning. A number of reports have suggested that mobile learning and mail can in fact help students to be more motivated and therefore improve their overall performance. Email-learning introduces a whole new type of learning, different from the typical tasks such as making notes, reading from a textbook, etc. This variety within a learning course is sure to increase levels of interest among learners [7]. The Online Learning Support Framework is models that can help educators understand the distinct needs of the online student. In phase two there is a considerable amount of teaching that needs to occur, along with a high level of relational support. Relational in this context means personalized support. Students do not require hand-holding; they are past the 'needy' phase but they do require feedback, reinforcement and acknowledgement for the time and energy they have already invested [8]. It is the personal connection and encouragement that is so very needed in online learning. It is the instructor's encouragement that can make the difference between whether a student is successful or not. Yet how can the instructor bridge the virtual divide of time and space to encourage and motivate. Respond to student questions within 24 hours [9]. Doing so encourages focus and commitment – the student senses that someone 'cares'. Provide timely feedback on assignments. At the beginning of phase two, the student should ideally receive feedback on at least one individual assignment. This provides the instructor with the opportunity to make a connection with the student, thus encourage and motivate a student just at the point when the student is moving towards becoming an independent learner. This reinforces the student's increasing level of self-direction. Comment strategically within discussion boards – making note of insightful or notable comments made by students. The advent of information technology in various fields has lead the large volumes of data storage in various formats like records, files, documents, images, sounds, scientific data and many new data formats [15].

4. IMPLEMENTATION PROCESS:

Abstraction Scheme: Here we introduced an E-mail & SMS abstraction scheme using IMAP, SMTP and ID3 algorithm. ID3 procedure is presented to depict the generation process of students.

4.1 IMAP:

Internet Message Access Protocol **(IMAP)** [16] is a protocol for e-mail retrieval and storage developed by Mark Crispin in 1986 at Stanford University as an alternative to POP.IMAP [16]. Some clients and servers preferentially use vendor-specific, proprietary_protocols, but most support SMTP for sending e-mail and POP and IMAP for retrieving [16] e-mail, allowing interoperability with other servers and clients [16].

The IMAP Algorithm

1.For each message:

A. If **id_table** contains an *empty* Container for this ID   Store this message in the Container's message slot.

Else:

- Create a new Container object holding this message;
- Index the Container by Message-ID in **id_table**

B. For each element in the message's References field:

- Find a Container object for the given Message-ID:

  - If there's one in **id_table** use that;
  - Otherwise, make (and index) one with a null Message.

2. Find the root set: Walk over the elements of **id_table**, and gather a list of the Container objects that have no parents.

1. **Discard id_table.**
2. **Prune empty containers.**
3. **Group root set by subject.**
4. **Now you're done threading!**
5. **Now, sort the siblings.**

At this point, the parent-child relationships are set. However, the sibling ordering has not been adjusted, so now is the time to walk the tree one last time and order the siblings by date, sender [16], subject, or whatever. This step could also be merged in to the end of step 4, above, but it's probably clearer to make it be a final pass. If you were careful, you could also sort the messages *first* and take care in the above algorithm to not perturb the ordering, but that doesn't really save anything [16].

4.2 SMTP:

Simple Mail Transfer Protocol (SMTP) [15] is an Internet standard for electronic mail (e-mail) transmission [15]. First defined by RFC 821 in 1982, it was last updated in 2008 with the Extended SMTP additions by RFC 5321 - which is the protocol in widespread use today. SMTP by default uses TCP port 25. The protocol for mail submission is the same, but uses port 587. While electronic mail servers and other mail transfer agents use SMTP to send and receive mail messages, user-level [15] client mail applications typically use SMTP only for sending messages to a mail server for relaying. For receiving messages, client applications usually use either POP3 or IMAP. While proprietary systems (such as Microsoft Exchange and Lotus Notes/Domino) and web mail systems (such as Hotmail, Gmail and Yahoo! Mail) use their own non-standard protocols [15]to access mail box accounts on their own mail servers, all use SMTP when sending or receiving email from outside their own systems.

The SMTP Algorithm

There are three steps to SMTP mail transactions [15]. The transaction is started with a MAIL command which gives the sender   identification. A series of one or more RCPT commands follows giving the receiver information.   Then a DATA command gives the mail data.   And finally, the end of mail data indicator confirms the transaction [15].

1. MAIL <SP> FROM :< reverse-path> <CRLF>

2. RCPT <SP> TO :< forward-path> <CRLF>
 This command gives a forward-path identifying one recipient.
- If accepted, the receiver-SMTP returns a 250 OK reply, and Stores the forward-path. If the recipient is unknown the
- Receiver-SMTP returns a 550 Failure reply. This second step of
- The procedure can be repeated any number of times

3. DATA <CRLF> the mail data includes the memo header items such as Date, Subject, To, Cc, From.

4.2 ID3 Algorithm:

ID3 is a nonincremental algorithm, meaning it derives its classes from a fixed set [11] of training instances. An incremental algorithm revises the current concept definition, if necessary, with a new sample. The classes created by ID3 are inductive, that is, given a small set [17]of training instances, the specific classes created by ID3 are expected to work for all future instances [11]. The distribution of the unknowns must be the same as the test cases. Induction classes cannot be proven to work in every case since they may classify an infinite number of instances. Note that ID3 (or any inductive algorithm) may misclassify data [11].

Data Description

The sample data used by ID3 has certain requirements, which are [11]:

Attribute-value [17] description - the same attributes must describe each example and have a fixed number of values.

Predefined classes - an example's attributes must already be defined, that is, they are not learned by ID3 [11].

Discrete classes - classes must be sharply delineated. Continuous classes broken up into vague categories such as a metal being "hard, quite hard, flexible, soft, quite soft" are suspect [17].

Sufficient examples – since [11] inductive generalization is used (i.e. not provable) there must be enough test cases to distinguish valid patterns from chance occurrences [17].

Attribute Selection:

How does ID3 decide which attribute is the best? A statistical property [11], called information gain, is used. Gain measures how well a given attribute separates training examples into targeted classes. The one with the highest information (information being the most useful for classification) is selected. In order to define gain, we first borrow an idea from information theory called entropy [11].

3354

Entropy measures the amount of information in an attribute.

Given a collection S of c outcomes

Entropy(S) = S -p(I) log2 p(I)

where p(I) is the proportion of S belonging to class I. S is over c. Log2 is log base 2.

Example:

If S is a collection of 14 examples with 9 YES and 5 NO examples then [11]

Entropy(S) = - (9/14) Log2 (9/14) - (5/14) Log2 (5/14) = 0.940

Notice entropy is 0 if all members of S belong to the same class (the data is perfectly classified). The range of entropy is 0 ("perfectly classified") to 1 ("totally random").

Gain(S, A) is information gain of example set S on attribute A is defined as

Gain(S, A) = Entropy(S) - S ((|Sv| / |S|) * Entropy(Sv)) [11]

Where:

S is each value v of all possible values of attribute A [11]

Sv = subset of S for which attribute A has value v

|Sv| = number of elements in Sv

|S| = number of elements in S

ID3 Algorithm:

ID3(Examples,Target Attribute, Attributes)

1. Create a root node for the tree

2. If all examples are positive, return the single-node

   Tree Root, with label=+

3. if all examples are negative return the single-node

   Tree Root, with label=-

4. If number of predicting attributes is empty, then

   Return the single node tree Root with label=most

   Common value of the target attribute in the example

5. Otherwise Begin

A. A=The Attribute that best classifies examples

B. Decision Tree attribute for Root=A

C. For each possible value, vi, of A

   a. Add a new tree branch below Root, Corresponding to the A=vi

   b. Let Examples(vi) be the subset of Examples that have the value vi for A

   c. If Examples (vi) is empty → Then below this new branch add a Leaf node with Label=most common target value in the Examples

   d. Else below this new branch add the subtree ID3

   Examples(vi)),Target_Attribute, Attributes-{A})

6. End

7. Return Root

5. RESULTS AND ANALYSIS: One classification rules can be generated for each path [11] from each terminal node to root node [11]. Pruning technique was executed by removing nodes with less than desired number of objects [17]. IF-THEN rules may be easier to understand. The expected student results are shown in below figure. Above variables were defined for the present investigation as follows [11]: **LP:** lab Performance
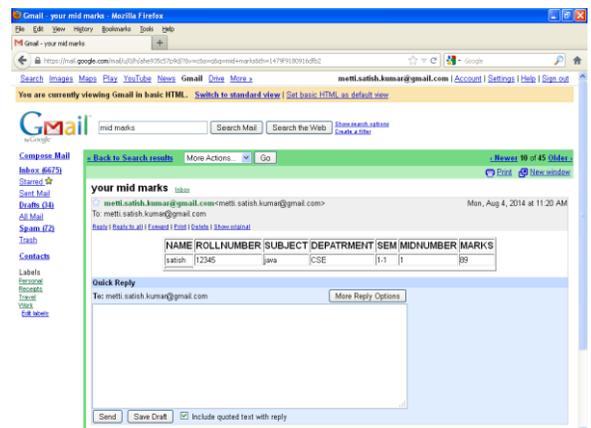

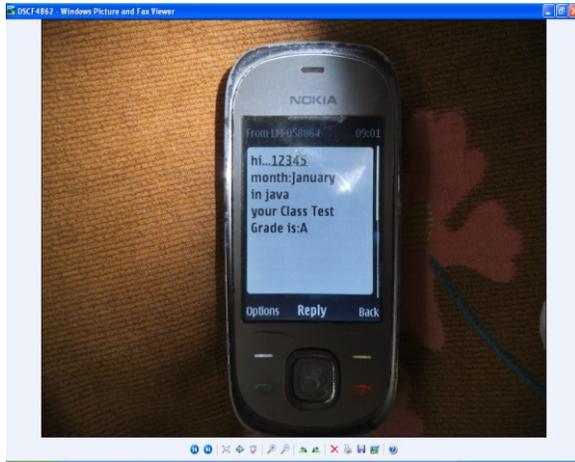Fig 1: Mid marks email received by student

3355

Fig 2: Class test grade SMS received by student mobile [14]

IF PSM = „First AND ATD = „Good" AND CTG = „Good or „Average THEN ESR = First
IF PSR = „First AND CTG = „Good AND ATD = "Good OR „Average THEN ESM = „First
IF PSR = „Second AND ATD = „Good" AND ASS = „Yes THEN ESR = „First
IF PSR = „Second AND CTG = „Average AND LP= „Yes THEN ESR = „Second
IF PSR = „Third AND CTG = „Good OR „Average AND ATD = "Good OR „Average THEN PSR = „Second
IF PSrR= „Third AND ASS = „No AND ATD = „Average THEN PSR = „Third
IF PS = „Fail AND CTG = „Poor AND ATD= „Poor THEN PSR = „Fail

Table 1: Rule Set generated by Decision Tree

Roll Num.	Subject	CTG	ATD	LP	ASS	PSR	ESR
1010	Java	A	Good	Poor	Poor	FRIST	FRIST CLASS
2020	Java	A	Poor	Poor	Poor	FRIST	FRIST CLASS
3030	Java	C	Avg	Poor	Poor	Third	SECOND CLASS
4040	Java	B	Avg	Avg	Poor	Second	SECOND CLASS
5050	Java	C	Poor	Poor	Poor	Third	THIRD CLASS
6060	Java	C	Avg	Poor	Avg	Third	THIRD CLASS
7070	Java	C	Poor	Poor	Poor	Fail	FAIL

Table 2: Predicate the End semester result of students

ASS: Assignment
PSR: Previous Semester Result
CTG: Class Test Grade
ESR: End Semester Result
ATD: attendance

6. CONCLUSION: Required information is provided to the students through mails and respective notifications through mobiles. For predicting the performance of the students at the end of the semester, student information like their attendance, and marks obtained in internal and external exams, assignments have been collected. By applying decision tree classification technique on the collected information the students have been categorized into different groups like excellent, good, average and poor. This study will be beneficial for the teachers to evaluate the performance of the students and from which they can take remedial action against the weak students to improve their performance. These results will be sent to students and their parents for analyzing their performance in the upcoming exams.

7. REFERENCES:

[1]. Mario Muñoz-Organero, Pedro J. Muñoz-Merino," Sending Learning Pills to Mobile Devices in Class to Enhance Student Performance and Motivationin Network Services Configuration Courses"Carlos Delgado Kloos," IEEE TRANSACTIONS ON EDUCATION, VOL. 55, NO. 1, 2012.

[2]. Rugarcia, R. M. Felder, D. R. Woods, and J. E. Stice, "The future of engineering education—I: A vision for a new century," *Chem. Eng. Educ., vol. 34, no. 1, pp. 16–25, 2000.*

[3].A. Holzinger, A. Nischelwitzer, and M. Meisenberger, "Mobile phones as a challenge for m-learning: Examples for Mobile Interactive Learning Objects (MILOs)," in *Proc. 3rd IEEE PerCom Workshops,* 2005, pp. 307–311.

[4].E. Syukur and S. W. Loke, "MHS learning services for pervasive campus environments," in *Proc. 5th IEEE PerCom Workshops, 2007,* pp. 204–210.

[5].H. Ogata, "Computer supported ubiquitous learning: Augmenting Learning experiences in the real world," in *Proc. 5th IEEE WMUTE,* Beijing, China, 2008, pp. 3–1

[6]Syvanen, R. Beale, M. Sharples, M. Ahonen, and P. Lonsdale, "Supporting pervasive learning environments: Adaptability and context wareness in mobile learning," in *Proc. IEEE WMTE*, Tokushima, Japan, 2005, pp. 251–253.

3356

[7]. K. Kim, "Ubiquitous learning supporting system for future classroom in Korea," in *Proc. Soc. Inf. Technol. Teacher Educ. Int. Conf.*, Mar. 2008, pp. 2648–2657.

[8]. K. Scott and R. Benlamri, "Context-aware services for smart learning spaces," *IEEE Trans. Learn. Technol.*, vol. 3, no. 3, pp. 214–227, Jul. –Sep. 2010.

[9]. K. Scott and R. Benlamri, "Context-aware services for smart learning
spaces," *IEEE Trans. Learn. Technol.*, vol. 3, no. 3, pp. 214–227, Jul. –Sep. 2010.

[10]. N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of
techniques for predicting academic performance," in *Proc. IEEE FIE
Conf.*, Milwaukee, WI, 2007, pp. T2G-7–T2G

[11]. Brijesh Kumar Baradwaj "
Mining Educational Data to Analyze Students" Performance," IJACSA Vol. 2, No. 6, 2011

[12]. R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation
of intrinsic motivation, social development, and well-being," *Amer. Psych.*, vol. 55, pp. 68–78, 2000.

[13]. J. M. Keller and C. M. Reigeluth, Eds., "Motivational design of instruction," in Instructional Design Theories and Models: An Overview of Their Current Status. Hillsdale, NJ: Erlbaum, 1983.

[14]. K. Swan, M. van't Hooft, A. Kratcoski, and D. Unger, "Uses and effects
of mobile computing devices in K–8 classrooms," *J. Res. Technol. Educ.*, vol. 38, no. 1, pp. 99–113, 2005.

[15]. M. Tariq Banday, "EFFECTIVENESS AND LIMITATIONS OF E-MAIL
SECURITY PROTOCOLS". , Vol.2, No.3, May 2011
[16]. Seyyed Hossein Raja, "Securing Risks of Electronic Mail Based on the Type of Organization" International Journal of Innovation, Management and Technology, Vol. 3, No. 3, June 2012
[17]. S. T. Hijazi, and R. S. M. M. Naqvi, Factors affecting student" s performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

About the authors:

Sateesh kumar Metti did his B.tech (Computer Science Engineering) in Sri vasavi engineering College, Pedatadepalli, West godavari district. Now he is purusing his M.Tech (Computer Science Engineering) in the same college from Jawaharlal Nehru Technological University, Kakinada. His area of interest is Data mining and Computer Networks.

R RajaRamesh Merugu is at present working as an Associate Professor in the Department of Computer Science and Engineering at Sri Vasavi Engineering College, Pedatadepalli, Tadepallegudem. He was Pursuing his PhD from Andhra University from 2010. He has Completed M.Tech in Computer Science Engineering from Andhra University, Vishakapatnam, 2004. He pursued MCA from University of Hyderabad, Hyderabad, 2001. He has a total teaching experience (UG and PG) of 13 years. His main areas of interest in Research include Software Engineering.